Improvement estimation on two successive occasions in cluster sampling

García Luengo, Amelia Victoria
Department of Mathematics. University of Almera, Almera, Spain.
amgarcia@ual.es

Shahzad, Usman
Department of Mathematics and Statistics PMAS-Arid Agriculture University, Rawalpindi, Pakistan.
usman.stat@yahoo.com

Koyuncu, Nursel
Department of Statistics University of Hacettepe, Ankara, Turkey.
nkoyuncu@hacettepe.edu.tr

Muhammad, Hanif
Department of Mathematics and Statistics PMAS-Arid Agriculture University, Rawalpindi, Pakistan.
mhpuno@hotmail.com

Abstract

In this paper we attempt the problem of estimation of the sum of mean and the change of mean in mail surveys. This problem is conducted for current occasion in the context of cluster sampling over sampling on two successive occasions. The sampling units are clusters and the observations on the first occasion are regarded as ancillary information for the observations on the second or current occasion.

The results obtained are demonstrated with the help of an empirical study, which reveals that under certain condition, the cluster sampling on two occasions is more efficient than the simple random sampling on two occasions.

Keywords: Successive sampling, Cluster sampling, Sum of population mean, Change of population mean, Efficiency.

1. Introduction

In sample surveys cluster or area sampling is widely practised because of its low cost and time saving device to conduct large scale and complicated surveys. Its use becomes more desirable when a list of elements is not available or units of the population are widely scattered and it is required to take repeated observations on the selected units (Pradhan, 2004).

It is well known that the cluster sampling is better than the simple random sampling when the intra class correlation within the same cluster is negative and smaller than $-(M-1)^{-1}$, where M denotes the size of the cluster. The relative efficiency of the cluster sampling is controlled by both the size of the cluster and the intra-class

correlation coefficient, it decreases if the size of the cluster increases substantially (Sukhatme and Sukhatme, 1970). Zarkovich and Krane, 1965 demonstrated that the correlation between two characters in cluster sampling with clusters as sampling units is expected to be higher than correlation coefficient in element sampling.

Pradhan, 2004 and Singh and Kumar, 2011 have proposed estimators in the estimation of the current population mean under the set up of cluster sampling on two occasions. Sampling on successive occasions was first considered by Jessen, 1942 in the analysis of farm data and this theory was further extended by Patterson, 1950, Eckler et al., 1955, Rao and Graham, 1964, Gupta, 1979, Das, 1982, Singh and Singh, 2001, Artes Rodriguez and Garcia Luengo, 2005, Garcıa Luengo and Oña, 2010 among others.

Continuing this line of work, we develop the Hansen et al., 1953 technique to estimate the sum of mean and the change of mean for current occasion in the context of cluster sampling over sampling on two successive occasions. An empirical study that allows us to investigate the performance of the proposed strategy is carried out.

2. Notation

Suppose that the population is composed of N clusters of M elements each, and that a simple random sample of n clusters is drawn without replacement from it.

Let (x_{ij}, y_{ij}), $(i = 1, 2, .., N; j = 1, 2, .., M)$ be the values of the characteristic on first and second occasions for the jth unit of the ith cluster, respectively.

A simple random sample (without replacement) of n clusters is drawn on the first occasion. On the second occasion, a simple random sample of $m = n\lambda$ $(0 < \lambda < 1)$ clusters (i.e. Mm elements) of the n clusters selected on the first occasion is retained (matched) while an independent sample of $u = n\mu = n - m$ $(\mu = 1 - \lambda)$ clusters (i.e. Mu elements) is replaced (unmatched with the first occasion) from the entire population. The characters x and y are supposed to be correlated when they are observed on the same unit repeatedly.

Define,

$$\bar{X}_{i.} = \frac{1}{M} \sum_{j=1}^{M} x_{ij}; \quad \bar{Y}_{i.} = \frac{1}{M} \sum_{j=1}^{M} y_{ij}$$

The means of the ith cluster on the first and second occasion respectively.

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} \bar{x}_{i.}; \quad \bar{Y} = \frac{1}{N} \sum_{i=1}^{N} \bar{y}_{i.}$$

The cluster population mean of x and y respectively.

$$\bar{X}_{NM} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} x_{ij}; \quad \bar{Y}_{NM} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij}$$

The population means of x and y per element on the first and second occasions respectively.

$$S_x^2 = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} (x_{ij} - \bar{X}_{NM})^2}{NM - 1}; \quad S_y^2 = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} (y_{ij} - \bar{Y}_{NM})^2}{NM - 1}$$

The population mean square between elements on the first and second occasions respectively.

$$\rho_x = \frac{\sum\limits_{i=1}^{N} \sum\limits_{j<k}^{M} (x_{ij} - \bar{X}_{NM})(x_{ik} - \bar{X}_{NM})}{(M-1)(NM-1)S_x^2}; \quad \rho_y = \frac{\sum\limits_{i=1}^{N} \sum\limits_{j<k}^{M} (y_{ij} - \bar{Y}_{NM})(y_{ik} - \bar{Y}_{NM})}{(M-1)(NM-1)S_y^2}$$

The intra-class correlation coefficient between elements of a cluster on first and second occasions respectively.

$$\rho_b = \frac{\sum\limits_{i=1}^{N} (\bar{Y}_{i.} - \bar{Y}_{NM})(\bar{X}_{i.} - \bar{X}_{NM})}{\left[\sum\limits_{i=1}^{N} (\bar{Y}_{i.} - \bar{Y}_{NM})^2 \sum\limits_{i=1}^{N} (\bar{X}_{i.} - \bar{X}_{NM})^2 \right]^{1/2}}$$

The simple correlation coefficient between cluster means on both occasions.

$$\bar{x}_{nM} = \frac{1}{nM} \sum_{l=1}^{nM} x_l; \quad \bar{y}_{nM} = \frac{1}{nM} \sum_{l=1}^{nM} y_l$$

Sample means based on a simple random sample of nM units.

$$\bar{x}_{uM} = \frac{1}{uM} \sum_{l=1}^{uM} x_l; \quad \bar{y}_{uM} = \frac{1}{uM} \sum_{l=1}^{uM} y_l$$

Sample means based on a simple random sample of uM units.

$$\bar{x}_{mM} = \frac{1}{mM} \sum_{l=1}^{mM} x_l; \quad \bar{y}_{mM} = \frac{1}{mM} \sum_{l=1}^{mM} y_l$$

Sample means based on a simple random sample of mM units.

3. Estimation of the sum of mean in cluster sampling on two successive occasions

Consider the following minimum variance linear unbiased estimator of the sum

$$z_1 = a\,\bar{x}_{uM} + b\,\bar{x}_{mM} + c\,\bar{y}_{mM} + d\,\bar{y}_{uM} \tag{1}$$

which expected value is given by

$$
\begin{aligned}
\mathrm{E}(z_1) &= a\mathrm{E}(\bar{x}_{uM}) + b\mathrm{E}(\bar{x}_{mM}) + c\mathrm{E}(\bar{y}_{mM}) + d\mathrm{E}(\bar{y}_{uM}) \\
&= a\bar{X} + b\bar{X} + c\bar{Y} + d\bar{Y} = (a+b)\bar{X} + (c+d)\bar{Y} = \bar{X} + \bar{Y}
\end{aligned}
$$

Unbiasedness of z_1 implies $a + b = 1$ and $c + d = 1$, so that $b = 1 - a$ and $d = 1 - c$. Substituting the value of b and d in equation (1), we obtain

$$
z_1 = a\,\bar{x}_{uM} + (1-a)\bar{x}_{mM} + c\,\bar{y}_{mM} + (1-c)\,\bar{y}_{uM} \tag{2}
$$

The variance of z_1 is given by

$$
\mathrm{V}(z_1) = a^2\mathrm{V}(\bar{x}_{uM}) + (1-a)^2\,\mathrm{V}(\bar{x}_{mM}) + c^2\,\mathrm{V}(\bar{y}_{mM}) + (1-c)^2\,\mathrm{V}(\bar{y}_{uM}) + 2(1-a)\,c\,\mathrm{Cov}(\bar{x}_{mM}, \bar{y}_{mM}) \tag{3}
$$

Assuming that N is sufficiently large, other covariance terms being zero, the variances and covariance involved in (3) are given by

$$
\mathrm{V}(\bar{x}_{uM}) = \frac{1}{uM}\{\tilde{\rho}_x S_x^2\}; \quad \mathrm{V}(\bar{x}_{mM}) = \frac{1}{mM}\{\tilde{\rho}_x S_x^2\}
$$

$$
\mathrm{V}(\bar{y}_{uM}) = \frac{1}{uM}\{\tilde{\rho}_y S_y^2\}; \quad \mathrm{V}(\bar{y}_{mM}) = \frac{1}{mM}\{\tilde{\rho}_y S_y^2\}
$$

where

$$
\tilde{\rho}_x = 1 + \rho_x(M - 1); \quad \tilde{\rho}_y = 1 + \rho_y(M - 1); \quad \mathrm{Cov}(\bar{x}_{mM}, \bar{y}_{mM}) = \frac{1}{mM}\rho_b\sqrt{\tilde{\rho}_x\tilde{\rho}_y}S_x S_y
$$

Minimizing the variance of z_1 with respect to a and c, the optimum values of a and c are:

$$
a = \frac{\mu(1 - \mu\rho_b^2)}{1 - \mu^2\rho_b^2} + \frac{\lambda\mu\rho_b}{1 - \mu^2\rho_b^2}\frac{S_y}{S_x}\sqrt{\frac{\tilde{\rho}_y}{\tilde{\rho}_x}}; \qquad c = \frac{\lambda}{1 - \mu^2\rho_b^2} - \frac{\lambda\mu\rho_b}{1 - \mu^2\rho_b^2}\frac{S_x}{S_y}\sqrt{\frac{\tilde{\rho}_x}{\tilde{\rho}_y}}
$$

Using the optimum values of a and c, the estimator z_1 reduces to

$$
z_1 = \frac{\mu(1 - \mu\rho_b^2)}{1 - \mu^2\rho_b^2}(\bar{x}_{uM} + \bar{y}_{uM}) + \frac{\lambda}{1 - \mu^2\rho_b^2}(\bar{x}_{mM} + \bar{y}_{mM})
$$

$$
+ \frac{\lambda\mu\rho_b}{1 - \mu^2\rho_b^2}\left[(\bar{x}_{uM} - \bar{x}_{mM})\frac{S_y}{S_x}\sqrt{\frac{\tilde{\rho}_y}{\tilde{\rho}_x}} + (\bar{y}_{uM} - \bar{y}_{mM})\frac{S_x}{S_y}\sqrt{\frac{\tilde{\rho}_x}{\tilde{\rho}_y}}\;\right]
$$

In case $S_x = S_y$ and $\tilde{\rho}_x = \tilde{\rho}_y$, z_1 reduces to

$$
z_1 = \frac{\lambda}{1 + \mu\rho_b}(\bar{x}_{mM} + \bar{y}_{mM}) + \frac{\mu(1 + \rho_b)}{1 + \mu\rho_b}(\bar{x}_{uM} + \bar{y}_{uM})
$$

1040

Pak.j.stat.oper.res. Vol.XV No.IV 2019 pp1037-1050

with variance

$$\mathrm{V}(z_1) = \frac{2S_y^2 \tilde{\rho}_y (1 + \rho_b)}{nM(1 + \mu\rho_b)} = (1 + \rho_y(M-1)) \frac{(1 + \rho_b)}{(1 + \mu\rho_b)} \frac{2S_y^2}{nM} \tag{4}$$

We note that, for $\rho_b > 0$, equation (4) is minimum for $\mu = 0$, i.e., the variance of z_1 is minimized if the clusters on both occasions are independent. In this case,

$$\mathrm{V}(z_1) = \frac{2S_y^2 \tilde{\rho}_y (1 + \rho_b)}{nM} = (1 + \rho_y(M-1)) \frac{2S_y^2 (1 + \rho_b)}{nM}$$

3.1 Efficiency of cluster sampling on two occasions

If the samples on both occasions are drawn using SRSWOR, the variance of the optimum estimator z neglecting the finite population correction factor is given by

$$\mathrm{V}(z) = \frac{(1 + \rho)}{(1 + \mu\rho)} \frac{2S_y^2}{nM}$$

where ρ is the simple correlation coefficient between values of units on first and second occasion. The relative efficiency of z_1 compared to z is

$$\frac{\mathrm{V}(z)}{\mathrm{V}(z_1)} = \frac{(1 + \rho)(1 + \mu\rho_b)}{(1 + \rho_y(M-1))(1 + \mu\rho)(1 + \rho_b)}$$

The cluster sampling on both occasions provides more efficient estimate than the simple random sampling on both occasions if

$$\rho_y \leqslant \frac{1}{M-1} \left[\frac{(\mu - 1)(\rho_b - \rho)}{(1 + \mu\rho)(1 + \rho_b)} \right]$$

Further, in order that z_1 would be more efficient than z if

$$M \leqslant \frac{1}{\rho_y} \left[\frac{(\mu - 1)(\rho_b - \rho)}{(1 + \mu\rho)(1 + \rho_b)} \right] + 1$$

which gives the upper limit of M.
Tables: 1-3 have been computed below to show the relative efficiency of cluster sampling in sampling on two occasions compared to simple random sampling of elements for some specified values of ρ_y, ρ, ρ_b, M and μ.

Table 1: Relative efficiency of cluster sampling over simple random sampling

			$M = 2, \rho_y = 0.01, \mu = 0.5$				
ρ_b/ρ	0.5	0.6	0.7	0.8	0.9	0.95	0.98
0.1	1.1341	1.1632	1.1901	1.2151	1.2384	1.2494	1.2559
0.2	1.0891	1.1170	1.1429	1.1669	1.1893	1.1999	1.2061
0.3	1.0510	1.0780	1.1029	1.1261	1.1477	1.1579	1.1639
0.4	1.0184	1.0445	1.0687	1.0911	1.1120	1.1220	1.1277
0.5	0.9901	1.0155	1.0390	1.0608	1.0811	1.0908	1.0964

			$M = 4, \rho_y = 0.01, \mu = 0.5$				
ρ_b/ρ	0.5	0.6	0.7	0.8	0.9	0.95	0.98
0.1	1.1121	1.1406	1.1670	1.1915	1.2144	1.2252	1.2315
0.2	1.0680	1.0953	1.1207	1.1442	1.1662	1.1766	1.1826
0.3	1.0306	1.0570	1.0815	1.1042	1.1254	1.1354	1.1413
0.4	0.9986	1.0242	1.0479	1.0699	1.0904	1.1002	1.1058
0.5	0.9709	0.9958	1.0188	1.0402	1.0601	1.0696	1.0751

			$M = 5, \rho_y = 0.01, \mu = 0.5$				
ρ_b/ρ	0.5	0.6	0.7	0.8	0.9	0.95	0.98
0.1	1.1014	1.1296	1.1558	1.1801	1.2027	1.2134	1.2197
0.2	1.0577	1.0848	1.1099	1.1332	1.1550	1.1653	1.1713
0.3	1.0207	1.0469	1.0711	1.0936	1.1146	1.1245	1.1303
0.4	0.9890	1.0144	1.0379	1.0597	1.0800	1.0896	1.0952
0.5	0.9615	0.9862	1.0090	1.0302	1.0500	1.0593	1.0648

Table 2: Relative efficiency of cluster sampling over simple random sampling

			$M = 2, \rho_y = 0.05, \mu = 0.5$				
ρ_b/ρ	0.5	0.6	0.7	0.8	0.9	0.95	0.98
0.1	1.0909	1.1189	1.1448	1.1688	1.1912	1.2018	1.2081
0.2	1.0476	1.0745	1.0994	1.1224	1.1440	1.1542	1.1601
0.3	1.0110	1.0369	1.0609	1.0832	1.1040	1.1138	1.1196
0.4	0.9796	1.0047	1.0280	1.0496	1.0697	1.0792	1.0848
0.5	0.9524	0.9768	0.9994	1.0204	1.0400	1.0492	1.0547

4. Estimation of the change of mean in cluster sampling on two successive occasions

Consider the following minimum variance linear unbiased estimator of the change

$M = 4, \rho_y = 0.05, \mu = 0.5$						
ρ_b/ρ	0.5	0.6	0.7	0.8	0.9	0.95	0.98
0.1	0.9960	1.0216	1.0452	1.0672	1.0876	1.0973	1.1030
0.2	0.9565	0.9810	1.0038	1.0248	1.0445	1.0538	1.0592
0.3	0.9231	0.9467	0.9687	0.9890	1.0080	1.0169	1.0222
0.4	0.8944	0.9173	0.9386	0.9583	0.9767	0.9854	0.9905
0.5	0.8696	0.8919	0.9125	0.9317	0.9495	0.9580	0.9629

$M = 5, \rho_y = 0.05, \mu = 0.5$						
ρ_b/ρ	0.5	0.6	0.7	0.8	0.9	0.95	0.98
0.1	0.9545	0.9790	1.0017	1.0227	1.0423	1.0516	1.0570
0.2	0.9167	0.9402	0.9619	0.9821	1.0010	1.0099	1.0151
0.3	0.8846	0.9073	0.9283	0.9478	0.9660	0.9746	0.9796
0.4	0.8571	0.8791	0.8995	0.9184	0.9360	0.9443	0.9492
0.5	0.8333	0.8547	0.8745	0.8929	0.9100	0.9181	0.9228

Table 3: Relative efficiency of cluster sampling over simple random sampling

$M = 2, \rho_y = 0.1, \mu = 0.5$						
ρ_b/ρ	0.5	0.6	0.7	0.8	0.9	0.95	0.98
0.1	1.0413	1.0680	1.0927	1.1157	1.1371	1.1472	1.1531
0.2	1.0000	1.0256	1.0494	1.0714	1.0920	1.1017	1.1074
0.3	0.9650	0.9898	1.0127	1.0340	1.0538	1.0632	1.0687
0.4	0.9351	0.9590	0.9812	1.0019	1.0210	1.0302	1.0355
0.5	0.9091	0.9324	0.9540	0.9740	0.9927	1.0015	1.0067

$M = 3, \rho_y = 0.1, \mu = 0.5$						
ρ_b/ρ	0.5	0.6	0.7	0.8	0.9	0.95	0.98
0.1	0.9545	0.9790	1.0017	1.0227	1.0423	1.0516	1.0570
0.2	0.9167	0.9402	0.9619	0.9821	1.0010	1.0099	1.0151
0.3	0.8846	0.9073	0.9283	0.9478	0.9660	0.9746	0.9796
0.4	0.8571	0.8791	0.8995	0.9184	0.9360	0.9443	0.9492
0.5	0.8333	0.8547	0.8745	0.8929	0.9100	0.9181	0.9228

$$\Delta_1 = a\,\bar{x}_{uM} + (1-a)\bar{x}_{mM} + c\,\bar{y}_{mM} + (1-c)\,\bar{y}_{uM} \tag{5}$$

which expected value is given by

$$
\begin{aligned}
\mathrm{E}(\Delta_1) &= a\mathrm{E}(\bar{x}_{uM}) + b\mathrm{E}(\bar{x}_{mM}) + c\mathrm{E}(\bar{y}_{mM}) + d\mathrm{E}(\bar{y}_{uM}) \\
&= a\bar{X} + b\bar{X} + c\bar{Y} + d\bar{Y} = (a+b)\bar{X} + (c+d)\bar{Y} = \bar{Y} - \bar{X}
\end{aligned}
$$

ρ_b/ρ	$M = 5, \rho_y = 0.1, \mu = 0.5$						
	0.5	0.6	0.7	0.8	0.9	0.95	0.98
0.1	0.8182	0.8392	0.8586	0.8766	0.8934	0.9014	1.2857
0.2	0.7857	0.8059	0.8245	0.8418	0.8580	0.8656	1.1786
0.3	0.7582	0.7777	0.7957	0.8124	0.8280	0.8354	1.0879
0.4	0.7347	0.7535	0.7710	0.7872	0.8023	0.8094	1.0102
0.5	0.7143	0.7326	0.7496	0.7653	0.7800	0.7869	0.9429

Unbiasedness of Δ_1 implies $a + b = -1$ and $c + d = 1$, so that $b = -(a + 1)$ an $d = 1 - c$. Substituting the value of b and d in equation (5), we obtain

$$\Delta_1 = a\,\bar{x}_{uM} - (a+1)\bar{x}_{mM} + c\,\bar{y}_{mM} + (1-c)\,\bar{y}_{uM} \qquad (6)$$

The variance of Δ_1 is given by

$$V(\Delta_1) = a^2 V(\bar{x}_{uM}) + (a+1)^2\,V(\bar{x}_{mM}) + c^2\,V(\bar{y}_{mM}) + (1-c)^2\,V(\bar{y}_{uM}) - 2(a+1)\,c\,\mathrm{Cov}(\bar{x}_{mM}, \bar{y}_{mM}) \qquad (7)$$

Assuming that N is sufficiently large, other covariance terms being zero, the variances and covariance involved in (7) are given by

$$V(\bar{x}_{uM}) = \frac{1}{uM}\{\tilde{\rho}_x S_x^2\}; \quad V(\bar{x}_{mM}) = \frac{1}{mM}\{\tilde{\rho}_x S_x^2\}$$

$$V(\bar{y}_{uM}) = \frac{1}{uM}\{\tilde{\rho}_y S_y^2\}; \quad V(\bar{y}_{mM}) = \frac{1}{mM}\{\tilde{\rho}_y S_y^2\}$$

where

$$\tilde{\rho}_x = 1 + \rho_x(M - 1); \quad \tilde{\rho}_y = 1 + \rho_y(M - 1); \quad \mathrm{Cov}(\bar{x}_{mM}, \bar{y}_{mM}) = \frac{1}{mM}\rho_b\sqrt{\tilde{\rho}_x\tilde{\rho}_y}S_x S_y$$

We wish to choose whose values of a and c that minimize $V(\Delta_1)$. Equating the derivatives of $V(\Delta_1)$ with respect to a and c to zero, it follows that the optimum values are:

$$a = \frac{-\mu(1 - \mu\rho_b^2)}{1 - \mu^2\rho_b^2} + \frac{\lambda\mu\rho_b}{1 - \mu^2\rho_b^2}\frac{S_y}{S_x}\sqrt{\frac{\tilde{\rho}_y}{\tilde{\rho}_x}} \qquad c = \frac{\lambda}{1 - \mu^2\rho_b^2} + \frac{\lambda\mu\rho_b}{1 - \mu^2\rho_b^2}\frac{S_x}{S_y}\sqrt{\frac{\tilde{\rho}_x}{\tilde{\rho}_y}}$$

Using the optimum values of a and c, the estimator Δ_1 reduces to

$$\Delta_1 = \frac{\mu(1 - \mu\rho_b^2)}{1 - \mu^2\rho_b^2}(\bar{y}_{uM} - \bar{x}_{uM}) + \frac{\lambda}{1 - \mu^2\rho_b^2}(\bar{y}_{mM} - \bar{x}_{mM})$$

$$+ \frac{\lambda\mu\rho_b}{1 - \mu^2\rho_b^2}\left[(\bar{x}_{uM} - \bar{x}_{mM})\frac{S_y}{S_x}\sqrt{\frac{\tilde{\rho}_y}{\tilde{\rho}_x}} - (\bar{y}_{uM} - \bar{y}_{mM})\frac{S_x}{S_y}\sqrt{\frac{\tilde{\rho}_x}{\tilde{\rho}_y}}\;\right]$$

In case $S_x = S_y$ and $\tilde{\rho}_x = \tilde{\rho}_y$, Δ_1 reduces to

$$\Delta_1 = \frac{\lambda}{1 - \mu\rho_b}(\bar{y}_{mM} - \bar{x}_{mM}) + \frac{\mu(1 - \rho_b)}{1 - \mu\rho_b}(\bar{y}_{uM} - \bar{x}_{uM})$$

with variance

$$V(\Delta_1) = \frac{2S_y^2 \tilde{\rho}_y (1 - \rho_b)}{nM(1 - \mu\rho_b)} = (1 + \rho_y(M - 1))\frac{(1 - \rho_b)}{(1 - \mu\rho_b)}\frac{2S_y^2}{nM} \tag{8}$$

We note that, for $\rho_b > 0$, equation (8) is minimum for $\mu = 0$, i.e., the variance of Δ_1 is minimized if the clusters on both occasions are identical. In this case,

$$V(\Delta_1) = \frac{2S_y^2 \tilde{\rho}_y (1 - \rho_b)}{nM} = (1 + \rho_y(M - 1))\frac{2S_y^2(1 - \rho_b)}{nM}$$

4.1 Efficiency of cluster sampling on two occasions

If the samples on both occasions are drawn using SRSWOR, the variance of the optimum estimator Δ neglecting the finite population correction factor is given by

$$V(\Delta) = \frac{(1 - \rho)}{(1 - \mu\rho)}\frac{2S_y^2}{nM}$$

where ρ is the simple correlation coefficient between values of units on first and second occasion. The relative efficiency of Δ_1 compared to Δ is

$$\frac{V(\Delta)}{V(\Delta_1)} = \frac{(1 - \rho)(1 - \mu\rho_b)}{(1 + \rho_y(M - 1))(1 - \mu\rho)(1 - \rho_b)}$$

The cluster sampling on both occasions provides more efficient estimate than the simple random sampling on both occasions if

$$\rho_y \leqslant \frac{1}{M - 1}\left[\frac{(\mu - 1)(\rho - \rho_b)}{(\mu\rho - 1)(\rho_b - 1)}\right]$$

Further, in order that Δ_1 would be more efficient than Δ if

$$M \leqslant 1 - \frac{1}{\rho_y}\left[\frac{(\mu - 1)(\rho_b - \rho)}{(\mu\rho - 1)(\rho_b - 1)}\right]$$

which gives the upper limit of M.

Tables: 4-6 have been computed below to show the relative efficiency of cluster sampling in sampling on two occasions compared to simple random sampling of elements for some specified values of ρ_y, ρ, ρ_b, M and μ.

Table 4: Relative efficiency of cluster sampling over simple random sampling

$M = 2, \rho_y = 0.01, \mu = 0.5$				
ρ_b/ρ	0.1	0.2	0.3	0.4	0.5
0.5	1.4070	0.9001	0.9327	0.9626	0.9901
0.55	1.5112	1.4179	1.3137	1.1964	1.0634
0.6	1.6415	1.5402	1.4269	1.2995	1.1551
0.65	1.8090	1.6973	1.5725	1.4321	1.2730
0.7	2.0323	1.9069	1.7666	1.6089	1.4301
0.75	2.3450	2.2002	2.0384	1.8564	1.6502
0.78	2.6008	2.4402	2.2608	2.0590	1.8302

$M = 4, \rho_y = 0.01, \mu = 0.5$				
ρ_b/ρ	0.1	0.2	0.3	0.4	0.5
0.5	1.3797	0.8826	0.9146	0.9439	0.9709
0.55	1.4819	1.3904	1.2882	1.1731	1.0428
0.6	1.6096	1.5102	1.3992	1.2743	1.1327
0.65	1.7739	1.6644	1.5420	1.4043	1.2483
0.7	1.9928	1.8698	1.7323	1.5777	1.4024
0.75	2.2994	2.1575	1.9989	1.8204	1.6181
0.78	2.5503	2.3929	2.2169	2.0190	1.7946

$M = 5, \rho_y = 0.01, \mu = 0.5$				
ρ_b/ρ	0.1	0.2	0.3	0.4	0.5
0.5	1.3664	0.8741	0.9058	0.9348	0.9615
0.55	1.4676	1.3770	1.2758	1.1619	1.0328
0.6	1.5941	1.4957	1.3857	1.2620	1.1218
0.65	1.7568	1.6484	1.5271	1.3908	1.2363
0.7	1.9737	1.8519	1.7157	1.5625	1.3889
0.75	2.2773	2.1368	1.9796	1.8029	1.6026
0.78	2.5258	2.3699	2.1956	1.9996	1.7774

Table 5: Relative efficiency of cluster sampling over simple random sampling

$M = 2, \rho_y = 0.05, \mu = 0.5$

ρ_b/ρ	0.1	0.2	0.3	0.4	0.5
0.5	1.3533	0.8658	0.8971	0.9259	0.9523
0.55	1.4536	1.3639	1.2636	1.1507	1.0229
0.6	1.5789	1.4815	1.3725	1.2500	1.1111
0.65	1.7401	1.6327	1.5126	1.3776	1.2245
0.7	1.9549	1.8342	1.6993	1.5476	1.3757
0.75	2.2556	2.1164	1.9608	1.7857	1.5873
0.78	2.5017	2.3473	2.1747	1.9805	1.7605

$M = 4, \rho_y = 0.05, \mu = 0.5$

ρ_b/ρ	0.1	0.2	0.3	0.4	0.5
0.5	1.2357	0.7905	0.8192	0.8454	0.8696
0.55	1.3272	1.2453	1.1537	1.0507	0.9340
0.6	1.4416	1.3527	1.2532	1.1413	1.0145
0.65	1.5888	1.4907	1.3811	1.2578	1.1180
0.7	1.7849	1.6747	1.5516	1.4130	1.2560
0.75	2.0595	1.9324	1.7903	1.6304	1.4493
0.78	2.2842	2.1432	1.9856	1.8083	1.6074

$M = 5, \rho_y = 0.05, \mu = 0.5$

ρ_b/ρ	0.1	0.2	0.3	0.4	0.5
0.5	1.1842	0.7576	0.7850	0.8102	0.8333
0.55	1.2719	1.1934	1.1057	1.0069	0.8951
0.6	1.3816	1.2963	1.2010	1.0938	0.9722
0.65	1.5226	1.4286	1.3235	1.2054	1.0714
0.7	1.7105	1.6049	1.4869	1.3542	1.2037
0.75	1.9737	1.8519	1.7157	1.5625	1.3889
0.78	2.1890	2.0539	1.9029	1.7330	1.5404

Table 6: Relative efficiency of cluster sampling over simple random sampling

$M = 2, \rho_y = 0.1, \mu = 0.5$				
ρ_b/ρ	0.1	0.2	0.3	0.4	0.5
0.5	1,2919	0,8264	0,8564	0,8838	0,9091
0.55	1,3876	1,3019	1,2062	1,0985	0,9764
0.6	1,5072	1,4141	1,3102	1,1932	1,0606
0.65	1,6610	1,5584	1,4439	1,3149	1,1688
0.7	1,8660	1,7508	1,6221	1,4773	1,3131
0.75	2,1531	2,0202	1,8717	1,7045	1,5152
0.78	2,3880	2,2406	2,0758	1,8905	1,6804

$M = 4, \rho_y = 0.1, \mu = 0.5$				
ρ_b/ρ	0.1	0.2	0.3	0.4	0.5
0.5	1,0931	0,6993	0,7246	0,7479	0,7692
0.55	1,1741	1,1016	1,0206	0,9295	0,8262
0.6	1,2753	1,1966	1,1086	1,0096	0,8974
0.65	1,4054	1,3187	1,2217	1,1126	0,9890
0.7	1,5789	1,4815	1,3725	1,2500	1,1111
0.75	1,8219	1,7094	1,5837	1,4423	1,2821
0.78	2,0206	1,8959	1,7565	1,5997	1,4219

$M = 5, \rho_y = 0.1, \mu = 0.5$				
ρ_b/ρ	0.1	0.2	0.3	0.4	0.5
0.5	1,0150	0,6494	0,6729	0,6944	0,7143
0.55	1,0902	1,0229	0,9477	0,8631	0,7672
0.6	1,1842	1,1111	1,0294	0,9375	0,8333
0.65	1,3050	1,2245	1,1345	1,0332	0,9184
0.7	1,4662	1,3757	1,2745	1,1607	1,0317
0.75	1,6917	1,5873	1,4706	1,3393	1,1905
0.78	1,8763	1,7605	1,6310	1,4854	1,3203

5. Conclusions

In sampling on two occasions we have considered the estimation of the sum of mean and the change of mean for current occasion when the sampling units are clusters and the observations on the first occasion are regarded as ancillary information for the observations on the second or current occasion. Under certain condition, the cluster sampling on two occasions is more efficient than the simple random sampling on two occasions.

The obtained results have revealed that for fixed ρ_y (intra-class correlation coefficient) and ρ_b (correlation coefficient between cluster means) the efficiency increases with large increase in ρ ($\rho > \rho_b$) (the simple correlation coefficient between values of units on first and second occasion) for the estimation of the sum of mean and the efficiency

increases with large increase in ρ_b $(\rho_b > \rho)$ for the estimation of the change of mean. Further, for fixed ρ_b and ρ, the efficiency decreases with increase in ρ_y.

References

1. Artes Rodriguez, E. M. and Garcia Luengo, A. V. (2005). Multivariate indirect methods of estimation in successive sampling. *J Ind. Soc. Agril. Statist*, 59(2):97–103.

2. Das, A. (1982). Estimation of population ratio on two occasions. *Journal of the Indian Society of Agricultural Statistics*, 34(2):1–9.

3. Eckler, A. R. et al. (1955). Rotation sampling. *The annals of mathematical statistics*, 26(4):664–685.

4. García Luengo, A. V. and Oña, I. (2010). The non-response in the change of mean and the sum of mean for current occasion in sampling on two occasions. *Chilean Journal of Statistics*, 2(1):63–84.

5. Gupta, P. (1979). Sampling on two successive occasions. *Journal of Statistical Research*, 13(13):7–16.

6. Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). *Sample survey methods and theory. V. 1. Methods and applications. V. 2. Theory*. Number 519.52 H249. Wiley.

7. Jessen, R. J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Research Bulletin (Iowa Agriculture and Home Economics Experiment Station)*, 26(304):1.

8. Patterson, H. (1950). Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(2):241–255.

9. Pradhan, B. K. (2004). On efficiency of cluster sampling on sampling on two occasions. *Statistica*, 64(1):183–191.

10. Rao, J. and Graham, J. E. (1964). Rotation designs for sampling on repeated occasions. *Journal of the American Statistical Association*, 59(306):492–509.

11. Singh, G. and Singh, V. (2001). On the use of auxiliary information in successive sampling. *Journal of the Indian Society of Agricultural Statistics*, 54(1):1–12.

12. Singh, H. P. and Kumar, S. (2011). Subsampling the non-respondents in cluster sampling on sampling on two successive occasions. *Statistics in Transition new series*, 12(1):9–24.

13. Sukhatme, P. and Sukhatme, B. (1970). *Sampling theory of surveys with applications.* Food and Agriculture Organization, Rome, Second edition.

14. Zarkovich, S. and Krane, J. (1965). Some efficient ways of cluster sampling. *Proceedings of 35th session of International Statistical Institute, Belgrade.*